Vision Research 59 (2012) 64–71

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Integration time for the perception of depth from motion parallax

Mark Nawrot [a,*], Keith Stroyan [b]

[a] Center for Visual Neuroscience, Department of Psychology, North Dakota State University, Fargo, ND 58108, USA
[b] Math Department, University of Iowa, Iowa City, IA 52242, USA

ARTICLE INFO

Article history:
Received 12 July 2011
Received in revised form 26 January 2012
Available online 1 March 2012

Keywords:
Depth perception
Motion parallax
Pursuit eye movements
Binocular stereopsis
Motion masking

ABSTRACT

The perception of depth from relative motion is believed to be a slow process that "builds-up" over a period of observation. However, in the case of motion parallax, the potential accuracy of the depth estimate suffers as the observer translates during the viewing period. Our recent quantitative model for the perception of depth from motion parallax proposes that relative object depth (d) can be determined from retinal image motion ($d\theta/dt$), pursuit eye movement ($d\alpha/dt$), and fixation distance (f) by the formula: $d/f \approx d\theta/d\alpha$. Given the model's dynamics, it is important to know the integration time required by the visual system to recover $d\alpha$ and $d\theta$, and then estimate d. Knowing the minimum integration time reveals the incumbent error in this process. A depth-phase discrimination task was used to determine the time necessary to perceive depth-sign from motion parallax. Observers remained stationary and viewed a briefly translating random-dot motion parallax stimulus. Stimulus duration varied between trials. Fixation on the translating stimulus was monitored and enforced with an eye-tracker. The study found that relative depth discrimination can be performed with presentations as brief as 16.6 ms, with only two stimulus frames providing both retinal image motion and the stimulus window motion for pursuit (mean range = 16.6–33.2 ms). This was found for conditions in which, prior to stimulus presentation, the eye was engaged in ongoing pursuit or the eye was stationary. A large high-contrast masking stimulus disrupted depth-discrimination for stimulus presentations less than 70–75 ms in both pursuit and stationary conditions. This interval might be linked to ocular-following response eye-movement latencies. We conclude that neural mechanisms serving depth from motion parallax generate a depth estimate much more quickly than previously believed. We propose that additional sluggishness might be due to the visual system's attempt to determine the maximum $d\theta/d\alpha$ ratio for a selection of points on a complicated stimulus.

1. Introduction

Quick and successful locomotion relies on a moving observer's ability to accurately determine the relative positions of objects in the environment. Motion parallax, the shift in the relative position of these objects created by the observer's own movement, is a crucial visual cue for this task. The perception of depth from motion parallax relies upon the changing visual scene from the moving observer's viewpoint. As the observer moves, fixation is maintained on one particular object for some interval while objects nearer than this fixation point appear to move in one direction and more distant objects appear to move in the opposite direction. This relative movement of objects at different distances from the observer's point of fixation is the proximal retinal cue for motion parallax. An important goal for vision science is to determine how the visual system uses this pattern of retinal image motion to recreate an accurate perception of relative depth for objects in the scene.

The study of depth from motion is complicated by the fact that there are different geometric mechanisms of retinal motion depending on the direction of observer translation and movement of the head and eyes. Motion parallax typically refers to lateral observer translation with the head fixed, observing a rigid scene. This is the case where the combination of smooth eye pursuit and retinal image motion is important (Nawrot & Joyce, 2006). In contrast, the extensive study in "optic flow" usually involves some forward observer translation that gives rise to a different geometry. For example, when translation is straight toward the fixate the resulting retinal flow is expansion from a point, but when observer translation is purely lateral, perpendicular to the line of sight, there is no focus of expansion in the retinal image motion. Recognition of this fundamental dichotomy goes as far back as von Kreis, writing in von Helmholtz (1910), vol. III, Note 4, p. 371), "...The changes of which he (von Helmholtz) speaks are such as the observer would notice if he advanced forward without changing the attitude of his head or his eyes especially. In reality the phenomena are complicated by the fact that, supposing our attention is attracted, not by some object moving along with us, but by stationary external

* Corresponding author.
E-mail address: mark.nawrot@ndsu.edu (M. Nawrot).

0042-6989/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2012.02.007

objects, we are invariably in the habit of keeping the eyes fastened for a brief space on some definite point, by turning them so as to counteract the effect of the forward motion of the body…" von Kries concludes with, "Now these apparent motions are just as useful as those described by Helmholtz for forming estimates of distance; and the probability is that both of them generally contribute to the result in some way, although it would be hard to say exactly how." While the visual system may process the visual cues from these different geometries in similar ways, by restricting our study to lateral observer translation and excluding the effects of forward translation, we are beginning to understand the neural mechanisms serving depth from of motion parallax and unravel von Kreis' dichotomy.

Echoing von Kreis' dichotomy, our mathematical model is restricted to lateral observer translation whereas several classic papers (such as Koenderink and van Doorn (1976) and Longuet-Higgins and Prazdny (1980)) are set up to cover general motion. However, many specific results in that literature in are undefined in the case of lateral translation. For example, the relative depth formula (Eq. 2.13) of Longuet-Higgins and Prazdny (1980) is undefined for lateral translation because $W = 0$ and W appears in a denominator. (W represents forward translation of the observer. Their vanishing point is "at infinity" in the case of lateral translation and the formulas that use it are undefined. The results on forward motion *are* interesting, but that is the other half of von Kreis' dichotomy.) We could cast our results in their general framework (see Nawrot & Stroyan, 2009, p. 1972; Stroyan & Nawrot, in press, Section 5), but our model simplifies the depth formula in our analysis and uses a direct representation of the retinal motion and smooth pursuit cues studied in our experiments.

More importantly, even in the lateral translation case, where the formulas do not depend on division by W (e.g., Nakayama & Loomis, 1974), the perception of relative depth in these analyses depends on perception of observer translational velocity. Indeed, much of the classic theory on depth from motion (Hanes, Keller, & McCollum, 2008; Longuet-Higgins & Prazdny, 1980; Nakayama & Loomis, 1974) specifically relies on observer translation speed in relative depth formulas. However, the source of this information is not well specified, and it is clearly not a direct visual cue. Moreover, the difficulty is that the same visual flow is also proposed to be a source of information about observer translation (e.g., "self-motion", Lappe, Bremmer, & van den Berg, 1999). The connection between this "observer velocity" approach and our current approach is explained in (Nawrot & Stroyan, 2009, Section 2.2).

In another classic approach (e.g., Domini & Caudek, 1999; Fermuller & Aloimonos, 1997; Gordon, 1965; Koenderink & van Doorn, 1976; Koenderink and van Doorn, 1987; Perrone & Stone, 1994) the geometry of optic flow is used to derive relative depth strictly from retinal information. This approach does not invoke observer translation velocity or an extra-retinal signal. All of the information is retinal. In a sense, Freeman and Fowler (2000) combine both approaches in experiments on perception of slant with and without pursuit. In the case of observation with pursuit, they write that, "…the visual system must rely on extra-retinal information about translation speed…" but do not give the mechanism for this perception nor use that quantity directly in their theoretical analysis, but rather seem to assume it replaces the fixed-eye retinal translation flow in their model. However, in their conclusion they write, "We have assumed throughout that motion-based slant perception involves separate estimates of relative motion…and (observer) translation speed." So, while they demonstrated the significance of an extra-retinal pursuit signal in the perception of slant, it is unclear what specific role pursuit might serve. Subsequent research (Naji & Freeman, 2004; Nawrot, 2003; Nawrot & Joyce, 2006) demonstrates the necessary role of pursuit in unambiguous perception of depth sign in motion parallax.

In contrast to these "observer velocity" and "retinal information" approaches, the novel feature of our theoretical approach (Nawrot & Stroyan, 2009; Stroyan & Nawrot, in press) is the elimination of a need to assume the observer directly perceives his translation velocity, but instead uses the ratio of one retinal (motion) and one non-retinal (pursuit) cue as the primary mathematical quantity of the analytical model. There is growing neurophysiological and psychophysical evidence that this combination of retinal image motion and a pursuit signal is necessary (Nadler, Angelaki, & DeAngelis, 2008) and sufficient (Kim, Angelaki, & DeAngelis, 2011; Moeeny & Cumming, 2011; Nadler et al., 2009; Nawrot & Stroyan, 2009) for the case of lateral observer translation. Our basic model for the perception of unambiguous scaled depth from motion parallax is as follows: Consider the schematic below (Fig. 1) describing one simple example condition that produces motion parallax. In this example, an observer translates rightward while maintaining gaze on the fixate (F) at viewing distance (f) along the line perpendicular to the line of translation. We assume the eye crosses this line at time $t = 0$. In (Fig. 1), a distractor (D) is shown more distant (d) than the fixate also along this line. As the observer translates, fixation on point F produces a change in gaze angle α at the rate $d\alpha/dt$ (or change $d\alpha$ in a small time dt). It appears that $d\alpha$ corresponds to an internal signal from the pursuit eye movement system (Nadler et al., 2009; Nawrot & Joyce, 2006). Observer translation causes the retinal position of point D, described by the angle θ, to change at the rate $d\theta/dt$ (or change $d\theta$ in a small time dt). This corresponds to the retinal image motion of D, moving in relation to the fixation point.

The precise geometric relation between the fixation distance (f), retinal image motion ($d\theta$), and pursuit eye movement ($d\alpha$) and relative depth (d) is given by the motion/pursuit law (Stroyan & Nawrot, in press):

$$d = \frac{d\theta}{d\alpha}\frac{1}{1 - d\theta/d\alpha} * f \tag{1}$$

at time $t = 0$, but this more complicated expression is approximately equal to the simple approximation of the Motion/Pursuit Ratio:

$$\frac{d}{f} \approx \frac{d\theta}{d\alpha} \quad \text{or} \quad d \approx \frac{d\theta}{d\alpha} f \tag{2}$$

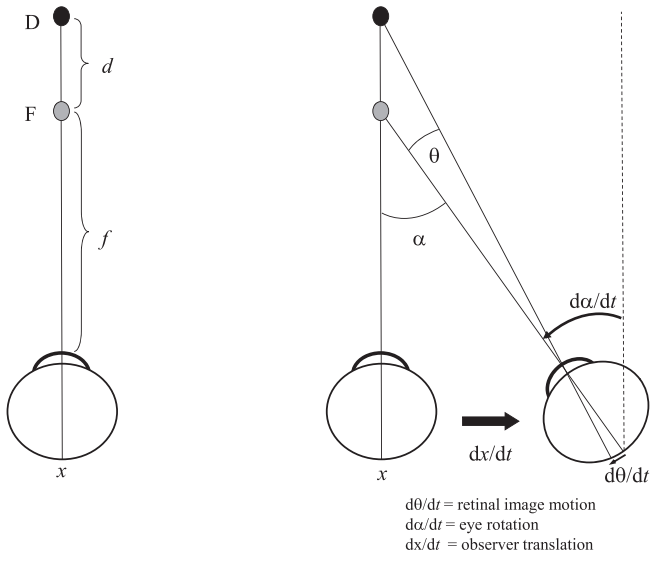

Fig. 1. Depiction of the geometry of motion parallax. The left panel depicts an eye viewing a point of fixation (F) at a viewing distance (f). The goal is to estimate the relative depth (d) of the distractor point (D). The right panel shows the two angles created by rightward observer translation at the rate dx/dt. The magnitude of pursuit corresponds the change in gaze angle α at the rate $d\alpha/dt$. The retinal motion of point D corresponds to the change in angle θ at the rate $d\theta/dt$.

at time $t = 0$ if the relative depth is small. For example, if the ratio is 0.1, then the difference between the M/P Law's calculation and the M/P Ratio's estimate of d is 10%.

Formulas (1) and (2) each account for both the direction and the magnitude of perceived relative depth of any distractor point. To expand on the case shown in Fig. 1, objects nearer and farther than the fixation point generate opposite directions of retinal image motion ($\pm d\theta$) and therefore generate estimates of d that are in opposite directions. A change in the direction of observer translation reverses the sign of both $d\theta$ and $d\alpha$, preserving the ratio, and giving constancy to the estimate of d. Changes in the point of fixation result in changes in f and $d\alpha$. Therefore, the model can generate different relative depth estimates (d) for each object while using the same singular values for the fixation distance (f) and pursuit ($d\alpha$) that reflects a specific point of fixation.

This time-varying nature of the motion/pursuit ratio is also important to the applicability of our model to perception. It is important to know the integration time (t) required by the human visual system to recover $d\theta/dt$, process $d\alpha/dt$, and then generate an estimate of relative depth (d). While these mathematical models use instantaneous retinal image and pursuit velocities, the human visual system requires a temporal interval to compute the velocities. But how long does that integration interval need to be?

The perception of depth from motion has been considered sluggish, and required to "build-up" over a period of observation (e.g., 750 ms, Eby, 1992; 800–1000 ms, Hildreth et al., 1990; 600–1000 ms, Andersen & Bradley, 1998; 700–1000 ms, Treue, Husain, & Andersen, 1991), especially for the perception of 3-D structure-from-motion such as when viewing a rotating kinetic depth figure (Wallach & O'Connell, 1953). If this processing sluggishness is true of motion parallax, then the potential accuracy of the depth estimates suffers as the translation of the observer changes the spatial relationship between the points, as viewed from the observer's vantage point. For instance, two equidistant objects are no longer equidistant as an observer moves in any direction other than directly between them. Moreover, even the viewing distance to the point of fixation changes as the observer translates. This effect of observer motion would not be an issue if the depth-from-motion process were instantaneous. But as this visual processing mechanism takes some interval of time, some amount of error or depth confusion becomes inherent to this depth perception process. Knowing these temporal limits will provide us an indication of the inherent depth perception limitations in motion parallax.

The current study seeks to determine the temporal parameters for the judgment of depth and motion of a motion parallax stimulus. Given the unusual challenges of studying temporal characteristics of perceptual processes, a high-contrast pattern-masking paradigm was used. That is, measuring the stimulus presentation duration required to perform a perceptual task is a poor measure of the underlying mechanism's temporal characteristics. This is because perceptual processing may continue after stimulus presentation has ceased. Similarly, reaction time measures are inappropriate because they rely on the characteristics of the motor system. For this reason the study of temporal characteristics of perceptual processes has given rise to a variety of masking paradigms (Breitmeyer, 1984). The rationale is that the subsequent presentation of a high-contrast pattern (the mask) can interfere with the perceptual processing of the immediately preceding stimulus. However, the interference effect of the masking stimulus has a limited time frame. If the perceptual processing is complete, the masking stimulus has no effect. The goal with a masking paradigm is to determine the stimulus presentation duration, or stimulus onset asynchrony (SOA), in which observer psychophysical performance is not affected by the mask. This is does not reveal the exact the duration of a visual process, but it does indicate a stage of completion to which the process is unaffected by the mask. That is, a masking paradigm allows the comparison of SOAs between different conditions and allows the conclusion that different processes have different temporal aspects. In the current study, the SOAs for masked and unmasked stimuli are compared for the conditions in which the observer must discriminate depth from motion, discriminate the direction of stimulus motion driving pursuit, and discriminate the direction of local relative motion within the stimulus. If the motion and depth conditions have similar SOAs, one might conclude that the temporal limitation on the depth perception process is found in motion processing. However, if motion and depth have different SOAs, one might conclude that depth requires additional processing, which we propose includes the integration of an internal pursuit signal to disambiguate the perception of depth from motion.

2. Methods

2.1. Apparatus

Stimuli were presented on a 21″ Sony Trinitron CRT with a 120 Hz refresh-rate (8.3 ms per frame) controlled by a Macintosh computer. Observers viewed from 57 cm with their head held in a headrest. Observers initiated trials and entered responses using the computer keyboard.

Eye position was monitored throughout the experiment with a video-based eye tracking system (Applied Science Laboratories, Bedford, MA) running under the control of a dedicated Windows-based computer. The eye tracking system sent monocular gaze position coordinates to the experimental computer through an analog connection from the eye tracking system. The experimental computer received the analog eye position coordinates using a 16-bit multifunction I/O board (National Instruments, Austin, TX). Prior to each block of trials, observers completed a 9-position calibration routine for the eye tracker control software. This was followed by a two-position gaze calibration for the experimental computer system.

2.2. Stimuli

In the current experiment stimuli translated across the monitor face while the observer remained stationary and pursued a fixation point at the center of the stimulus. This simplifies the motion parallax paradigm by requiring only pursuit eye movements, which correspond to $d\alpha$. In contrast, a motion parallax paradigm relying on observer head translation invokes a combination of both pursuit and translational vestibular ocular response eye movements. Only the pursuit signal is believed to be the extra-retinal signal involved in the perception of unambiguous depth from motion parallax (Nawrot & Joyce, 2006).

The design of motion parallax stimuli has been detailed previously (Nawrot & Joyce, 2006). In the current study motion parallax stimuli comprised 4000, 2×2 min dots within a stimulus window size of 6.6×6.6 deg. The stimulus window translated horizontally with a fixed velocity of 4.0 deg/s. The stimulus depicted 1.5 cycles of depth corrugation, with 0.75 cycles above and below the center point. To generate these depth corrugations, dots within the stimulus window translated horizontally in relation to the center (fixation) point with a sinusoidal velocity profile with a peak dot translation velocity (within the stimulus window) of 2.0 deg/s. Therefore the stimulus had a maximum M/P Ratio of ±0.5. The direction of dot translation ($d\theta$) in relation to direction of the stimulus window translation ($d\alpha$) determines the "depth phase" or the location of the "peaks" and "valleys" in the stimulus. That is, dots generating retinal image motion ($d\theta$) in the opposite direction (image motion in the same direction) as $d\alpha$ are perceived near in

depth, and depth phase can be varied independently of stimulus window translation (dα). The observer's task was to report the depth phase of the stimulus (e.g., whether the top or bottom part of the stimulus appeared nearer than the fixation point).

The high-contrast pattern-masking stimulus was composed of 50 black and 50 white circles of varying sizes drawn upon a large (13 deg $\times$ 13 deg) black background. The circles were drawn, in alternating contrast, largest (7 deg) to smallest (2 deg) in random position upon the background. For each block of trials several different exemplars of the masking stimulus were created and used.

2.3. Participants

Five naïve observers performed a 2AFC discrimination of the depth phase of common random-dot motion parallax stimuli. All had normal or corrected-to-normal acuity. One observer completed six conditions and was excluded from participating in the last two conditions due to pain medication prescribed subsequent to orthopedic surgery. The medication severely affected this observer's performance for one direction of eye movement, for an unusually long period of time, and her abnormal performance will be described in a subsequent report.

2.4. Procedure

The observer was seated with head supported in a chin rest. Each trial began with a fixation point at the center of the display. The observer initiated stimulus presentation with a key press. The fixation dot moved (laterally, left or right) immediately to the stimulus window starting position. Here the stimulus window would appear after a brief interval, randomly varying between 750 and 1500 ms. The stimulus window starting point was calculated so that the stimulus would appear moving across the center of the monitor, leftward moving stimuli started to the right of center, and right-ward moving stimuli started to the left of center. Stimuli having longer presentation durations started farther away from center.

The stimuli were presented in two interleaved staircases that varied the stimulus presentation duration for stimulus-window motion in leftward and rightward window movement. Both the leftward and rightward staircases started at 167 ms presentation duration and descended in 42 ms steps until its' first reversal or the 42 ms total presentation duration. All subsequent staircase steps were in 8.3 ms intervals. The staircase decision rule tracked to 79% correct (Wetherill & Levitt, 1965) with each of the interleaved staircases ending with six reversals, and the observer's performance determined from the average of the last four reversals.

Pursuit was monitored and enforced to a 1 deg window about the fixation point with the eye-tracking system. For each trial, stimulus presentation would only commence when the observer had established fixation on the stimulus fixation mark prior to initiating the trial. The 1 deg window was large enough, and stimulus translation speed low enough, to accommodate the initiation of pursuit without triggering a fixation loss. If a fixation loss was triggered anytime during a trial, then the stimulus was immediately extinguished, no observer response was recorded, and the trial was repeated. This enforcement procedure ensured that observers maintained fixation on, or very near, the fixation spot for each trial. Overall, only a small percentage of trials (<2–3%) were repeated.

2.4.1. Experimental conditions

Observers completed three of the interleaved staircases in each of the four experimental conditions, with and without the masking stimulus. In all four conditions observers reported perceived depth phase of the unambiguous motion parallax stimulus. The goal was to determine the minimum presentation interval required for an observer to accurately perform the depth phase (e.g., near–far) judgment. The pattern mask stimulus was used in "masking" conditions to determine whether the underlying visual processes required a longer presentation interval (SOA) than required in the non-masked condition. An initial pursuit interval (a pursuit prelude) was added before the presentation of the translating motion parallax stimulus to determine whether an ongoing pursuit signal would shorten the required presentation interval compared to conditions wherein observers programmed and initiated a pursuit eye movement concomitant with motion parallax stimulus presentation.

2.4.1.1. Condition 1: depth. In this condition observers viewed only the translating motion parallax stimulus and reported on the perceived depth phase. Here observers maintained their gaze on the fixation spot at the center of the translating stimulus as it moved leftwards or rightwards.

2.4.1.2. Condition 2: depth with mask. In this condition the translating motion parallax stimulus (identical to Condition 1) was followed (0 ms ISI) by the high-contrast pattern-masking stimulus. The pattern mask was centered upon the last fixation point and was about 4-times larger than the motion parallax stimulus. The pattern mask was presented for 83 ms.

2.4.1.3. Condition 3: depth with pursuit prelude. In this condition the translating motion parallax stimulus was presented after a brief, variable (750–1500 ms) pursuit prelude during which observers viewed only a translating fixation target. After the variable pursuit interval, the translating motion parallax stimulus was presented along with the pursuit target, with the motion parallax stimulus presentation being identical to that of Conditions 1 and 2 above.

2.4.1.4. Condition 4: depth with mask and pursuit prelude. This condition included both the mask (Condition 2) and pursuit prelude (Condition 3). Trials began with the variable duration pursuit prelude (750–1500 ms), continued with translating motion parallax stimulus (identical to Condition 1), and concluded with the pattern mask (0 ms ISI, 83 ms duration).

2.4.2. Control conditions

A set of control conditions allows comparisons between the time required to perform a judgment of motion direction and a judgment of depth phase. Although retinal image motion is often unappreciated by the observer in conditions of motion parallax – objects generating retinal image motion are often perceived as stationary, not moving (e.g., see Ono & Ujike, 2005) – motion provides one of the crucial input signals to the motion parallax process. Therefore, it is important to determine presentation duration necessary for the visual system to determine the motion characteristics of the stimulus, in the absence of depth. Similar temporal intervals for the perception of motion and depth would suggest that the temporal processing limitations are due to the motion processing mechanism alone. One might expect identical temporal intervals for motion and depth if the perception of depth from motion parallax were based solely on retinal image motion. However, if the perception of depth requires a longer temporal interval, this result would suggest that motion processing is not the processing stage causing the temporal limitation. Instead, this would suggest that the perception of depth requires some additional time-consuming processing, such as the generation of an internal pursuit eye movement signal.

2.4.2.1. Condition 5: stimulus relative motion. The goal of this condition was to determine the presentation duration required for observers to detect the directions of relative motion in the stimulus. In this case the stimulus window remained stationary in the

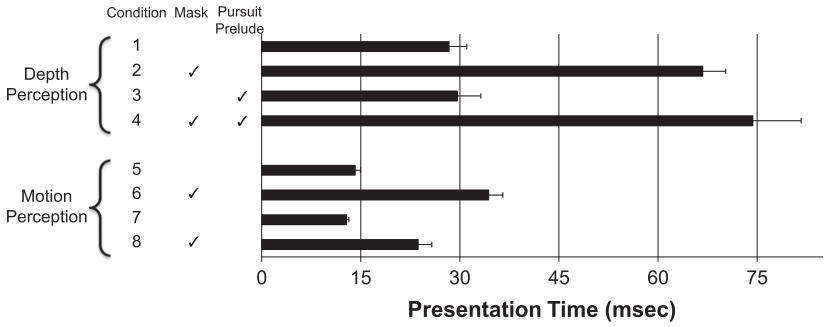

Fig. 2. Average presentation durations (ms) required to accurately perform depth (top) and motion (bottom) discriminations. Error bars denote one standard error.

center of the screen and observers performed judgments on the relative dot motion within the stimulus window. However, observers often reported perceiving some (ambiguous) depth in this stimulus, so the sine-wave modulation of dot motion was replaced by a square-wave modulation using the peak dot velocity of the sinusoidal stimulus (2.0 deg/s). Observers reported the perceived direction of stimulus motion in a particular portion of the stimulus. In this condition observers maintained fixation on the stationary fixation point.

2.4.2.2. Condition 6: stimulus relative motion with mask. In this condition the relative motion stimulus (identical to Condition 5) was followed (ISI 0 ms) with the pattern mask stimulus used in the experimental Conditions 2 and 4 above. Again, the presentation duration of the pattern mask was 83 ms.

2.4.2.3. Condition 7: stimulus window motion. In this condition the stimulus window, containing the relative motion stimulus from Condition 5, translated leftward or rightward. Here the stimulus window started from the center position so that the starting point would not provide a cue to the direction of window movement. Here observers maintained fixation on the point at the center of the translating stimulus window.

2.4.2.4. Condition 8: stimulus window motion with mask. In this condition the standard pattern mask (0 ms ISI, 83 ms duration) was presented subsequent to the translating relative motion stimulus presentation (used in Condition 7). Again, observers maintained fixation on the point at the center of the translating stimulus window.

3. Results

For each staircase a presentation duration threshold was calculated from the average of the last four reversals. For each condition an average threshold was calculated from the three different thresholds for each observer. Thresholds for leftward and rightward window motion were compared to determine whether observers showed a temporal difference for naso-temporal and temporo-nasal directions of eye movements. We found no evidence of a difference in normal observers. For example, the thresholds in the motion parallax (Condition 1, no masking stimulus, no pursuit prelude) showed only a small difference between rightward (29.6 (0.49) ms) and leftward (27.2 (0.47) ms) window movement directions that was not statistically significant ($p = 0.157$). Therefore, thresholds for leftward and rightward window motion were combined to provide a single mean threshold for each condition. These means for each condition are shown in Fig. 2.

3.1. Depth from motion parallax (experimental conditions)

The perception of unambiguous depth from motion parallax can be achieved with very brief stimulus presentations of about 30 ms (Condition 1). Adding a 750–1500 ms pursuit "lead-in" did not affect the presentation duration required to correctly perceive depth from motion parallax (Condition 3). The presentation duration remained at about 30 ms. However, the presentation of a pattern mask had a large effect on the required presentation duration, regardless of whether the trials had a pursuit prelude (Condition 4) or not (Condition 2). The average presentation duration required to perform the depth phase discrimination was increased to 67 ms and 74 ms for these two conditions respectively.

3.2. Motion perception (control conditions)

Observers can accurately report the direction of stimulus relative motion (Condition 5) and window motion (Condition 7) with very brief presentations of around 15 ms. Given the fixed frame rate of 120 Hz (8.3 ms per frame) this brief presentation duration corresponds to observers having little trouble performing these motion judgments with only 2 stimulus frames. The pattern mask did affect the presentation interval required to correctly perceive the direction of motion. It took about 35 ms to correctly judge the direction of relative dot motion within the stimulus (Condition 6) and about 25 ms to correctly judge the direction of stimulus window motion (Condition 8). These intervals required to determine motion direction are very similar to the intervals reported by Bischof and Di Lollo (1995) in their classic motion masking study.

4. Conclusion

A brief 30 ms presentation can be sufficient for an observer to unambiguously perceive depth from motion parallax. This indicates that the neural mechanisms serving the perception of depth from motion parallax are unexpectedly brisk compared to the "sluggish" estimates of close to a second of presentation time for depth from motion to "build up" (Andersen & Bradley, 1998; Eby, 1992; Hildreth et al., 1990; Treue, Husain, & Andersen, 1991). Such a brief integration time (≈30 ms) suggests that the ultimate precision of depth from motion parallax is unlikely to be limited predominately by the temporal properties of the neural processes. For example, consider that in 30 ms, a walking observer (5 km/h) translates less than 5 cm. In this case, the true relative distance between two inline objects at 1 m and 1.25 m changes by less than a millimeter. However, in this example, observer translation over a period of 0.5 s would change the relative distance of the two objects by several centimeters. Given this change in the quantity being estimated over this long temporal interval, it is difficult to

imagine how the visual system could ever generate an accurate estimate if the perceptual "build-up" time were close to one second.

The M/P Law provides a precise mathematical framework within which to consider such implications of the current results. Consider the example given above, for two in-line objects at 1 m and 1.25 m, using the M/P Law (Formula (2)) the estimate of depth changes only from the true relative distance of 0.25 m at $t = 0$ to an estimate of 0.249 m at $t = 0.03$ s. If the 30 ms translation is centered on the point when the objects are in-line ($t = -0.015$ to $t = +0.015$) the depth estimate generates M/PL is between 0.2498 and 0.25 (see Supplementary material for an interactive computation).[1] In contrast, the M/P Ratio (Formula (1)), which has the perceptual foreshortening of depth, generates relative depth estimates between 0.1999 m and 0.2 m during this 30 ms time interval for this particular viewing condition.

The reader can perform similar numerical examples for themselves (with a free browser plug-in) using interactive programs published at the Wolfram Demonstration Site, specifically Stroyan (2008a, 2008b). In the demonstration, one point can be made the fixate (1 m) and the other the distractor (1.25 m). Using the M/P Law (Formula (2)), and M/P Ratio (Formula (1)) this interactive demonstration shows how the relative depth between the fixate and the distractor changes as the observer translates. The reader can compare the estimate of d at two points of the translation by varying time, t, on the demonstration slider. These two demonstrations (Stroyan, 2008a, 2008b) assume a translation speed of 6.5 cm/s, an approximate interocular translation distance in one second. This translation speed is slower than the examples above, but provides a useful way to understand how the relative depth between objects changes with observer translation. The crucial point of the demonstration is that a 30 ms temporal integration does not suggest a significant limitation to the potential accuracy of the motion/pursuit cue for depth perception.

While the perception of depth from motion parallax can be accurately recovered from a brief 30 ms presentation, the underlying neural processing could actually take a longer duration to complete. A better estimate of this is provided by the pattern mask conditions. In these masking conditions the required presentation duration for accurate depth perception is about 70–75 ms. This means that a high-contrast pattern mask can interrupt performance of the depth-phase discrimination task if presented sooner than 75 ms after stimulus presentation begins.

Is this because the pattern mask affects the perception of motion, which in turn affects the perception of depth from motion? The control conditions (Conditions 5–8) indicate that this is not the explanation. In masking conditions, the direction of relative stimulus motion ($d\theta$) and the direction of the stimulus window movement (the proximal signal used by the visual system to generate $d\alpha$) can be accurately perceived with 15 ms presentations and 30 ms presentations. These results mean that the processing of visual motion has progressed beyond the effect of the pattern mask by about 30 ms. This is a much shorter presentation interval than that required to accurately perceive depth from motion parallax under the identical masking conditions. This means that the visual system has acquired both the translational and relative motion signals within the 30 ms presentation. However, this is insufficient for the system to generate the depth percept. This result is inconsistent with the idea that the perception of depth from motion parallax is based solely on motion processing. Additional processing, which can still be disrupted by the high-contrast pattern mask for the next 40–45 ms, must be occurring in the visual system.

We propose that this 70–75 ms interval is linked to the time course of generating the internal eye-movement signal ($d\alpha$) necessary for the unambiguous perception of depth from motion parallax (Nadler et al., 2009; Nawrot & Joyce, 2006; Nawrot & Stroyan, 2009). If we consider information flow along the pursuit eye movement pathway (see Kowler, 2011; Krauzlis, 2005 for recent reviews), motion signals are processed quickly (30 ms), but additional stages such as attention (Bisley & Goldberg, 2003), target selection (Ferrera & Lisberger, 1997), and programming the eye-movement signal require additional time, leading to the large latencies to generate eye-movements. For instance, the ocular following response (OFR; Miles, 1998) has latencies in the <85 ms range, and pursuit initiation is typically 100–150 ms (Krauzlis, 2004, or even faster, Ferrera & Lisberger, 1995). Therefore, the 70–75 ms of processing time needed for this internal eye-movement signal to withstand disruption by the high-contrast pattern mask appears to be within the time period for the initiation of the OFR or pursuit.

Moreover, the high-contrast pattern mask could be disrupting the internal pursuit eye-movement signal in a variety of ways. One possible explanation is simple oculomotor capture (Ludwig, Ranson, & Gilchrist, 2008; Theeuwes et al., 1998) where an abrupt stimulus onset (like the masking stimulus here) interferes with oculomotor programming. That is, the onset of the masking stimulus could "capture attention and evoke involuntary eye movements" (Lovejoy, Fowler, & Krauzlis, 2009) that compete with the pursuit signal processing. Another possible explanation is suggested by Keller and Khan's (1986) finding that a textured background (like the masking stimulus here) interferes with pursuit initiation, not just pursuit maintenance. Indeed, Kimmig, Miles, and Schwarz (1992) interpreted Keller and Khan's results as, "Thus neither the eyes nor the retinal images of the background need to move for textured backgrounds to impair subsequent pursuit,...". Therefore, it is possible that the masking stimulus in the present study impairs pursuit in a similar way.

However, it was unexpected that the pursuit prelude did not affect the 70–75 ms presentation SOA required to the high-contrast pattern mask (compare Condition 4 to Condition 2). Here the pursuit system was active for several hundred milliseconds prior to presentation of the translating motion parallax system, but the masking stimulus was still effective in disrupting the perception of depth for SOAs shorter than 70–75 ms. Knowing that the oculomotor system has a "velocity storage" mechanism, the hypothetical basis of anticipatory pursuit and other pursuit characteristics (Barnes & Collins, 2008; Barnes & Donelan, 1999), one might have expected that ongoing pursuit, initiated prior to presentation of the motion parallax stimulus, might have shortened the presentation interval required for accurate depth perception, prior to the mask. Nevertheless, the results show that the pursuit prelude had no effect. This result might be interpreted as meaning that the visual system does not use a "stored" pursuit velocity signal, and instead the depth perception process must receive an "active" eye movement signal, generated in the same time frame as the relative retinal image motion. That is, it appears that the visual system recovers $d\alpha$ and $d\theta$ together, in the same time frame. This type of synchrony is a reasonable strategy for the visual system to prevent a perceptual error of having an "old" (and now incorrect in velocity or direction) stored eye-movement signal paired with a new retinal image motion.

Nevertheless, it is difficult to discount the alternative explanation that within 70–75 ms the high-contrast pattern mask disrupts

[1] The supplemental materials contain an interactive program that can be used with the *Wolfram CDF Player*, available free at: http://www.wolfram.com/products/player/. The ".cdf" file suffix represents a document type that stands for "computational document format." These documents need a "reader" similar to the Adobe Reader for ".pdf" documents. Once you download and install the Wolfram CDF Player, you will be able to perform interactive computations with the document, similar to the walking example outlined above. The supplemental program uses the specific parameters of our experiment such as the $f = 57$ cm fixate distance and our 4 cm/s translation speed. At $t = 100$ ms the M/P Law computation of a distractor with $d = 40$ only differs in the fourth digit from the maximal value at $t = 0$ (0.7017 vs. 0.7018).

the overall "depth processing" mechanism, without specifically disrupting the generation of the internal pursuit signal. The specific parameters of this internal eye movement signal are difficult to measure directly because the observer's eyes do not actually have to move to disambiguate depth from motion parallax. That is, the internal eye movement signal used for motion parallax is separable from a measureable eye movement. There are several demonstrations that the internal eye movement signal, generated to countermand a reflexive eye movement signal like TVOR (Nawrot, 2003; Nawrot & Joyce, 2006) and OKR (Nawrot & Stockert, 2005) and thereby maintain stable fixation, is sufficient to disambiguate motion parallax. Of course, in these cases the pursuit signal is in the direction opposite the TVOR or OKR, so the perceived depth is reversed (e.g. see Fig. 11 in Nawrot & Joyce, 2006). All of these results show that an easily measureable external eye movement will not serve as a reliable proxy for the crucial internal eye movement signal.

One remaining question is: How does one reconcile the current study's demonstration of very short presentations needed to correctly report unambiguous depth-sign from a motion parallax stimulus with the results of previous studies suggesting that the perception of structure from motion is slow, and builds up over time (Andersen & Bradley, 1998; Eby, 1992; Hildreth et al., 1990; Treue, Husain, & Andersen, 1991)? A recent analysis by Stroyan & Nawrot (2011) suggests an explanation based on the dynamic nature of the M/P Ratio.

In the current experiment the corrugated motion parallax stimulus was constant over the horizontal axis, thereby making the fixate (*F*) and the relevant distractors (*D*, the maximum points of the peak and trough here) aligned along the naso-occipito axis. These conditions are well described by Fig. 1 and the M/P Law when $t = 0$ (Nawrot & Stroyan, 2009).

However, stimuli in the previous studies, in which depth was found to "build-up", had a complex 3-D shape that varied along the horizontal axis. The analysis of depth in such cases requires a more involved analysis (Stroyan & Nawrot, 2011) in which distractor points vary from the fixate in two dimensions. In the general case, the distractor (*D*) is displaced off the perpendicular line through the fixate (Fig. 3). However, in these cases, the strict mathematical relation in Formula (1) does not hold at $t = 0$. This is because the M/P Law has a time-varying characteristic that depends on relationship of *D* to the fixation point.

The motion/pursuit ratio, $d\theta/d\alpha[t]$, is a time-varying quantity that only gives a geometrically accurate cue to depth for certain points in space at certain times. Stroyan & Nawrot (2011) show that the maximum values of the motion/pursuit ratio can be geometrically aggregated to give an accurate representation of a 2-D shape (their Fig. 21). Their Fig. 20 shows the regions of space where this maximum is reached in three time intervals. The center panel shows a one second interval. Points in that part of space are accurately represented geometrically within that second, but points outside the region for the 30 ms interval on the left of Fig. 20 take longer until the observer reaches a position where the cue is geometrically accurate. Their Fig. 21 also shows the inaccuracy of the $t = 0$ cue for a 2-D shape (a circle with a diameter with the $t = 0$ cue on the left and the accurate aggregation on the right.) Therefore, it appears that the mathematically derived increase in the M/P Ratio and the psychophysically determined build-up in the perception of depth from motion may share a very similar time course, depending on the location of objects in the scene.

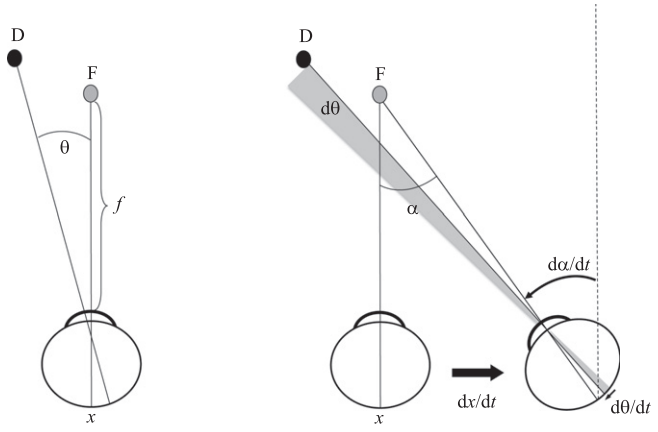

Fig. 3. Depiction of the geometry of motion parallax with the distractor displaced leftward from the line-of-sight position in the left panel. In this condition, the peak motion/pursuit ratio occurs when the hypothetical observer has moved to the right of the central position.

Perhaps a few words of explanation are in order to explain why the maximum motion/pursuit cue could take longer for a complicated shape, although Stroyan and Nawrot's (2011) results are geometrical and not empirical. The maximum of the motion/pursuit ratio occurs at the observer location where the relative depth is maximal. (This observation was not pointed out in Stroyan and Nawrot (2011), but they give a formula for the time when $d\theta/d\alpha$ is maximal and they occur at the same time.) Consider an observer shown in Fig. 3 at $t = 0$ on the left and a later time $t > 0$ on the right. If $\rho_D[t]$ is the straight line (radial) distance from the eye to the distractor at time t and $\rho_F[t]$ is the distance from the eye to the fixate, then $(\rho_D - \rho_F)/\rho_F$ is the relative depth of *D* beyond *F*. As the observer moves to the right from the $t = 0$ position, the depth of *D* beyond *F*, $(\rho_D - \rho_F)$, gets bigger until *D*, *F*, and the observer's eye are in line. However, the denominator term ρ_F in relative depth is also increasing, so tends to make the relative depth quantity decrease. The combined effect is that relative depth has a maximum at an observer position as shown in Fig. 3 on the right, after the $t = 0$ position, but before reaching the in-line position. This is the location where the motion/pursuit ratio for *D* is also maximal, giving the largest cue for *D*. As an observer moves, the "shape" of the 2D object builds up in time.

Therefore, one possible explanation for the longer integration times observed by (Hildreth, Grzywacz, Adelson, & Inada, 1990) and (Eby, 1992) is that the more complicated set-ups in their experiments forced their observers to use the time-varying properties of the motion/pursuit ratio described in (Stroyan & Nawrot, 2011). Their motions were different from ours, but if we consider a laterally translating observer (at speed 6.6 cm/s) with fixate at 100 cm and distractor 20 cm deeper and 3 cm to the left of the fixate (a distance of 20.2 cm), the motion/pursuit ratio will not reach it's maximum value until $t = 1.03$ s as shown on (Fig. 4).[2]

While the visual system is sensitive to differences in the M/P Ratio and perceives them as different depth magnitudes (Nawrot & Stroyan, 2009) it is unknown whether the visual system is sensitive to ongoing changes in M/P Ratio, and is able to detect an increase (build-up) and peak (end of a build-up) in the M/P Ratio. However, the similarity of the changes in the M/P Ratio to the build-up in the perception of depth from motion parallax suggests that the relationship between the two will be an important avenue of psychophysical investigation.

Regardless of how the perception of depth from motion might build-up or change during observer translation, the results of the

[2] The .cdf attachment in the supplemental materials (described above for computing the M/P Ratio) has a second part. For rough comparison to Eby (1992), we used a parabolic shape corresponding to the horizontal cross section of his stimulus. Our translation is linear, rather than circular, but about the same size – and the size can be varied interactively. The interactive program shows the times and positions where the maximal predictions are reached for a number of points on the arcs and tracks 6 of those points.

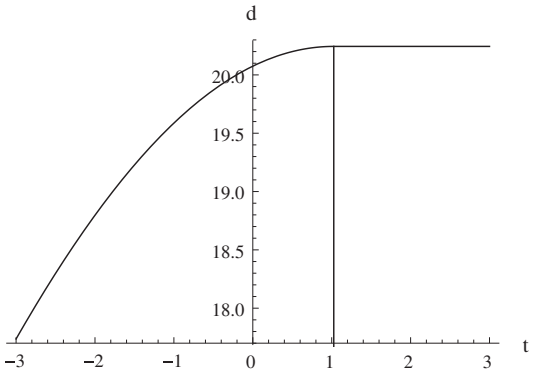

Fig. 4. The time-varying (t) motion/pursuit law prediction of structural depth (d) of an object 20 cm deeper and 3 cm to the left of the fixate at 100 cm.

current study show that depth from motion parallax can be perceived with brief presentations of about 30 ms duration. This means that motion parallax can provide a quick source of relative depth information for an observer moving through a cluttered environment. Knowing the lower temporal integration limit for motion parallax provides a starting point to determine the ultimate accuracy of depth from motion parallax for various viewing conditions.

Acknowledgments

This work was supported by a Centers of Biomedical Research Excellence (COBRE) Grant: NIH P20 RR020151 and P20 GM103505. The National Center for Research Resources (NCRR) is a component of the National Institutes of Health (NIH). The contents of this report are solely the responsibility of the authors and do not necessarily reflect the official views of the NIH or NCRR.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2012.02.007.

References

Andersen, R. A., & Bradley, D. C. (1998). Perception of three-dimensional structure from motion. *Trends in Cognitive Science, 2*, 222–228.

Barnes, G. R., & Collins, C. J. S. (2008). Evidence for a link between the extra-retinal component of random-onset pursuit and the anticipatory pursuit of predictable object motion. *Journal of Neurophysiology, 100*, 1135–1146.

Barnes, G. R., & Donelan, A. S. (1999). The remembered pursuit task: Evidence for segregation of timing and velocity storage in predictive oculomotor control. *Experimental Brain Research, 129*, 57–67.

Bischof, W. F., & Di Lollo, V. (1995). On motion and metacontrast with simultaneous onset of the stimuli. *Journal of the Optical Society of America A, 12*, 1623–1636.

Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral interparietal area and spatial attention. *Science, 299*, 81–86.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.

Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 426–444.

Eby, D. W. (1992). The spatial and temporal characteristics of perceiving 3-D structure from motion. *Perception & Psychophysics, 51*, 163–178.

Fermuller, C., & Aloimonos, Y. (1997). On the geometry of visual correspondence. *International Journal of Computer Vision, 21*, 223–247.

Ferrera, V. P., & Lisberger, S. G. (1995). Attention and target selection for smooth pursuit eye movements. *Journal of Neuroscience, 15*, 7472–7484.

Ferrera, V. P., & Lisberger, S. G. (1997). Neuronal responses in visual areas MT and MST during smooth pursuit target selection. *Journal of Neurophysiology, 78*, 1433–1446.

Freeman, T. C. A., & Fowler, T. A. (2000). Unequal retinal and extra-retinal motion signals produce different perceived slants of moving surfaces. *Vision Research, 40*, 1857–1868.

Gordon, D. A. (1965). Static and dynamic fields in human space perception. *Journal of the Optical Society of America, 55*, 1296–1303.

Hanes, D. A., Keller, J., & McCollum, G. (2008). Motion parallax contribution to perception of self-motion and depth. *Biological Cybernetics, 98*, 273–293.

Hildreth, E. C., Grzywacz, N. M., Adelson, E. H., & Inada, V. K. (1990). The perceptual buildup of three-dimensional structure from motion. *Perception & Psychophysics, 48*, 19–36.

Keller, E. L., & Khan, N. S. (1986). Smooth pursuit initiation in the presence of a textured background in monkey. *Vision Research, 26*, 943–955.

Kim, H. R., Angelaki, D. E., & DeAngelis, G. C. (2011). Neural correlates of depth perception from motion parallax in macaque area MT. Program No. 577.08. In *2011 Neuroscience meeting planner*. Washington, DC: Society for Neuroscience (online).

Kimmig, H. G., Miles, F. A., & Schwarz, U. (1992). Effects of stationary textured backgrounds on the initiation of pursuit eye movements in monkeys. *Journal of Neurophysiology, 68*, 2147–2164.

Koenderink, J. J., & van Doorn, A. J. (1976). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to the observer. *Optica Acta, 22*, 773–779.

Koenderink, J. J., & van Doorn, A. J. (1987). Facts on optic flow. *Biological Cybernetics, 56*, 247–254.

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research, 51*, 1457–1483.

Krauzlis, R. J. (2004). Recasting the smooth pursuit eye movement system. *Journal of Neurophysiology, 91*, 591–603.

Krauzlis, R. J. (2005). The control of voluntary eye movements: New perspectives. *Neuroscientist, 11*, 124–137.

Lappe, M., Bremmer, F., & van den Berg, A. V. (1999). Perception of self-motion from visual flow. *Trends in Cognitive Science, 3*, 329–336.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society London, 208*, 385–397.

Lovejoy, L. P., Fowler, G. A., & Krauzlis, R. J. (2009). Spatial allocation of attention during smooth pursuit eye movements. *Vision Research, 49*, 1275–1285.

Ludwig, C. J. H., Ranson, A., & Gilchrist, I. D. (2008). Oculomotor capture by transient events: A comparison of abrupt onsets, offsets, motion and flicker. *Journal of Vision, 8*, 1–16.

Miles, F. A. (1998). The neural processing of 3-D visual information: Evidence from eye movements. *European Journal of Neuroscience, 10*, 811–822.

Moeeny, A., & Cumming, B. (2011). Effect of pursuit eye movement on responses of MT neurons systematically depends on disparity and motion direction preference. Program No. 578.12. In *2011 neuroscience meeting planner*. Washington, DC: Society for Neuroscience (online).

Nadler, J. W., Angelaki, D. E., & DeAngelis, G. C. (2008). A neural representation of depth from motion parallax inmacaque visual cortex. *Nature, 452*, 642–645.

Nadler, J. W., Nawrot, M., Angelaki, D. E., & DeAngelis, G. C. (2009). MT neurons combine visual motion with a smooth eye movement signal to code depth sign from motion parallax. *Neuron, 63*, 523–532.

Naji, J. J., & Freeman, T. C. A. (2004). Perceiving depth order during pursuit eye movements. *Vision Research, 44*, 3025–3034.

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception, 3*, 63–80.

Nawrot, M. (2003). Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. *Vision Research, 43*, 1553–1562.

Nawrot, M., & Joyce, L. (2006). The pursuit theory of motion parallax. *Vision Research, 46*, 4709–4725.

Nawrot, M., & Stockert, C. (2005). Motion parallax in movies: Background motion, eye movement signals, and depth. *Journal of Vision, 5*(8), 644 (Abstract).

Nawrot, M., & Stroyan, K. (2009). The motion/pursuit law for visual depth perception from motion parallax. *Vision Research, 49*, 1969–1978.

Ono, H., & Ujike, H. (2005). Motion parallax driven by head movements: Conditions for visual stability, perceived depth, and perceived concomitant motion. *Perception, 34*, 477–490.

Perrone, J. A., & Stone, L. S. (1994). A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research, 43*, 2917–2938.

Stroyan, K. D., (2008a). *Motion pursuit law in 2D: Visual depth perception 3*. <http://demonstrations.wolfram.com/MotionPursuitLawIn2DVisualDepthPerception3/>.

Stroyan, K. D. (2008b). *Motion pursuit ratio and depth in 2D: Visual depth perception 18*. <http://demonstrations.wolfram.com/MotionPursuitRatioAndDepthIn2DVisualDepthPerception18/>.

Stroyan, K., & Nawrot, M. (in press). Visual depth from motion parallax and eye pursuit. *Journal of Mathematical Biology*. doi:10.1007/s00285-011-0445-1.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science, 9*, 379–385.

Treue, S., Husain, H., & Andersen, R. A. (1991). Human perception of structure from motion. *Vision Research, 31*, 59–75.

von Helmholtz, H. (1910/1925/1962). *Treatise on physiological optics*. New York: Dover (English translation by Southall JPC, three volumes bound as two, from the 3rd German edition of Handbuch der PhysiologischenOptik. von Kries Note 4 appeared in the 1925 edition).

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *The Journal of Experimental Psychology, 45*, 205–217.

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *The British Journal of Mathematical and Statistical Psychology, 18*, 1–10.